



US009183301B2

(12) **United States Patent**
Zwicky et al.

(10) **Patent No.:** **US 9,183,301 B2**
(45) **Date of Patent:** **Nov. 10, 2015**

(54) **SEARCH ENGINE OPTIMIZATION
PERFORMANCE VALUATION**

(71) Applicant: **Gere Dev. Applications, LLC,**
Wilmington, DE (US)

(72) Inventors: **Richard Kazimierz Zwicky,** Colwood
(CA); **Rafael Parra Hernandez,**
Victoria (CA)

(73) Assignee: **Gere Dev. Applications, LLC,**
Wilmington, DE (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 174 days.

(21) Appl. No.: **13/721,355**

(22) Filed: **Dec. 20, 2012**

(65) **Prior Publication Data**

US 2013/0185274 A1 Jul. 18, 2013

Related U.S. Application Data

(62) Division of application No. 12/555,224, filed on Sep.
8, 2009, now Pat. No. 8,364,529.

(60) Provisional application No. 61/094,821, filed on Sep.
5, 2008.

(51) **Int. Cl.**
G06F 17/30 (2006.01)
G06Q 30/02 (2012.01)

(52) **U.S. Cl.**
CPC **G06F 17/30864** (2013.01); **G06Q 30/02**
(2013.01); **G06Q 30/0244** (2013.01)

(58) **Field of Classification Search**
CPC G06F 17/30864; G06F 17/3053; G06F
17/30867; G06F 17/30; G06F 17/3089;
G06F 17/30424; G06F 17/30442; G06F 17/20;
G06F 17/3002; G06F 17/30554; G06F
17/30595; G06F 17/30646; G06F 17/30648;
G06F 17/30672; G06F 17/30705; G06F
17/3087; G06F 17/30876; G06F 17/30893;
G06F 17/30899; G06F 17/30991; G06F 19/24;
G06F 17/2264; G06F 17/30038; G06F
17/30312; G06F 17/30386; G06Q 30/02;
G06Q 30/0246; G06Q 30/0277; G06Q

30/0241; G06Q 30/0256; G06Q 30/08; G06Q
30/0201; G06Q 30/0202; G06Q 30/0242;
G06Q 30/0243; G06Q 30/0244; G06Q
30/0251; G06Q 30/0259; G06Q 30/0275;
G06Q 30/0625; G06Q 50/00; G06Q 30/0247;
G06Q 30/0254; G06Q 30/0633; G06Q
10/06395

USPC 705/14.45, 14.73, 14.43, 14.54, 14.42,
705/14.46, 26.3, 26.62, 7.31, 7.29, 14.41,
705/14.52, 26.8; 707/706, 723, E17.014,
707/E17.108, 709, 722, 728, 748, E17.017,
707/603, 710, 731, 737, 769, 812, 736, 740,
707/707, 756, 770; 709/217; 715/234

See application file for complete search history.

(56)

References Cited

U.S. PATENT DOCUMENTS

5,740,549 A 4/1998 Reilly et al.
6,006,260 A 12/1999 Barrick, Jr. et al.

(Continued)

OTHER PUBLICATIONS

Ramos, J. (Dec. 2003). Using tf-idf to determine word relevance in
document queries. In Proceedings of the First Instructional Confer-
ence on Machine Learning.*

(Continued)

Primary Examiner — Hosain Alam

Assistant Examiner — Diedra McQuitery

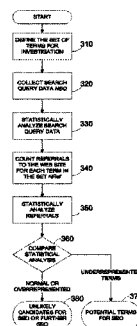
(74) *Attorney, Agent, or Firm* — Christensen O'Connor
Johnson Kindness PLLC

(57)

ABSTRACT

A way to promote a web site is via search engine optimization
(SEO). Traditionally, SEO practitioners have charged a fixed
fee for performing SEO services. An analyzer can determine
a first probability that a search query for a campaign term will
result in a referral from a search engine, determine second
probabilities that are associated with a plurality of particular
positions in a search engine results page, combine the search
query volume information, the first probability, and the sec-
ond probability for the position with a monetary value to
generate an organic price, combine the organic prices for the
plurality of positions to value the SEO services for the cam-
paign term. The analyzer evaluates the effectiveness of SEO
activities and can generate a volume-based value for SEO
services. In addition, disclosed techniques can also be used to
identify candidate terms for future SEO activities.

20 Claims, 5 Drawing Sheets



(56)

References Cited

U.S. PATENT DOCUMENTS

6,421,675	B1	7/2002	Ryan et al.	
6,714,917	B1	3/2004	Eldering et al.	
6,877,007	B1	4/2005	Hentzel et al.	
7,206,780	B2	4/2007	Slackman	
7,716,219	B2	5/2010	Sarukkai et al.	
7,809,605	B2 *	10/2010	Tonse et al.	705/14.54
8,001,131	B2	8/2011	Reese	
8,364,529	B1	1/2013	Zwicky et al.	
8,515,937	B1 *	8/2013	Sun et al.	707/707
8,769,079	B1 *	7/2014	Goswami	709/224
2002/0082923	A1	6/2002	Merriman et al.	
2002/0169760	A1	11/2002	Cheung et al.	
2003/0126013	A1	7/2003	Shand	
2004/0006459	A1 *	1/2004	Dehlinger et al.	704/10
2004/0107137	A1	6/2004	Skinner	
2004/0230491	A1	11/2004	Messer et al.	
2004/0267806	A1	12/2004	Lester	
2005/0076220	A1	4/2005	Zhang et al.	
2005/0097204	A1	5/2005	Horowitz et al.	
2005/0137939	A1 *	6/2005	Calabria et al.	705/26
2005/0222987	A1	10/2005	Vadon	
2006/0031206	A1	2/2006	Deubel et al.	
2006/0136294	A1	6/2006	Linden et al.	
2006/0136843	A1	6/2006	Shafron	
2006/0161520	A1	7/2006	Brewer et al.	
2006/0161534	A1	7/2006	Carson et al.	
2006/0161537	A1	7/2006	Amitay et al.	
2006/0161541	A1	7/2006	Cencini	
2006/0161542	A1	7/2006	Cucerzan et al.	
2006/0161591	A1	7/2006	Huang et al.	
2006/0161843	A1	7/2006	Ebrahimi et al.	
2006/0167852	A1	7/2006	Dissett et al.	
2006/0167854	A1	7/2006	Dissett et al.	
2006/0167857	A1	7/2006	Kraft et al.	
2006/0167874	A1	7/2006	Von Ahn Arellano et al.	
2006/0167896	A1	7/2006	Kapur et al.	
2006/0168056	A1	7/2006	Gandhi et al.	

2006/0200455	A1	9/2006	Wilson	
2006/0259360	A1	11/2006	Flinn et al.	
2006/0265435	A1	11/2006	Denissov	
2006/0265493	A1	11/2006	Brindley et al.	
2007/0005417	A1	1/2007	Desikan et al.	
2007/0073579	A1	3/2007	Immorlica et al.	
2007/0130131	A1	6/2007	Porter et al.	
2008/0162200	A1	7/2008	O'Sullivan et al.	
2008/0201235	A1	8/2008	Almeida	
2010/0036828	A1 *	2/2010	Carmel et al.	707/5
2010/0049695	A2	2/2010	Morsa	
2010/0205059	A1	8/2010	Yehoshua et al.	

OTHER PUBLICATIONS

Raleigh SEO Services Company, Search Engine Optimization NC, http://www.essex-seo.co.uk/archive/seo_services.html, Apr. 2009.*

10 Tactics to Rank Higher in Search Results, <http://www.wdfm.com/marketing-tips/rank-higher.php>, Nov. 2005.*

U.S. Appl. No. 12/143,387, filed Jun. 20, 2008, Caws.

Filip Radlinski & Thorsten Joachims, "Query Chain: Learning to Rank from Implicit Feedback", Proceedings of the ACM Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, Aug. 21-24, 2005, 10 pages.

Joachims, "Optimizing Search Engines using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining", Association for Computing Machinery, 2002, 10 pages.

New Search Patent Applications: Aug. 1, 2006, Microsoft Answers?; [online] [retrieved on Aug. 21, 2006] Retrieved from the Internet: <URL: <http://blog.searchenginewatch.com/060801232103>>, 6 pages.

Alex Cleanthous, "Estimating the Value of Search Engine Optimization," downloaded from <http://www.marketingpilgrim.com/2006/10/estimating-value-of-search-engine.html> on Aug. 10, 2009.

Andy Beal and Neal A. Rodriguez, "Bob and Weave Click Fraud Thieves", Oct. 9, 2006, www.marketingpilgrim.com/2006/10/bob-and-weave-click-fraud-thieves.html, 3 pages.

* cited by examiner

Web Images Maps News Shopping Gmail more

Sign in

Travel cuba

Search

Advanced Search
Preferences

Web

Results 1 - 10 of about 1,780,000 for Travel cuba (0.18 seconds)

Cuba Vacation Package

www.website1.com Cheap Vacation Packages to Cuba Call Today to Book a Trip and Save

Discount Cuba Vacations

www.website2.com Discount Cuba vacation packages with Canada's vacation specialists.

Cuba Special Deals

www.website3.com Save up to 60% on All Inclusive Packages Book online today!

USA CUBA TRAVEL : TRAVEL TO CUBA

A travel services company specializing in tours and holiday packages to Cuba lists information about destinations and current specials.

www.website4.com - Cached - Similar pages

Cuba travel Guide - Wikitravel

Open source travel guide to Cuba featuring up-to-date information on attractions, hotels, restaurants, nightlife, travel visas and more.

www.website5.com - Cached - Similar pages

Cuba Travel Guide and Travel Information - Lonely Planet

Lonely Planet Cuba Travel Guide. Cuba travel information, advice, tickets, reviews, maps and itineraries.

www.website6.com - Cached - Similar pages

Cuba Travel Guide - Cuba Vacations - Virtual Tourism

The best Cuba travel tips, reviews and photos, posted by real travelers. Research, plan and book your Cuba vacation with VirtualTourist.com.

www.website7.com - Cached - Similar pages

Cuba

Transactions related to tourist travel are not licensable. This restriction includes tourist travel To Cuba from or through a third country such as Mexico...

www.website8.com - Cached - Similar pages

Sponsored Links

Private Lodging in Cuba

Looking for a place particular?
Find your place to stay here!
www.website9.com

Last Minute Travel Deals

Latest specials on flights, hotels,
rentals & vacations. Great deals.
www.website10.com
British Columbia

Cuba Travel

Cuba Adventures for Canadians
Canada's Small Group Specialists
www.website11.com
British Columbia

Travel Cuba

Colonial Cities, Beaches & More
Expert Guides at Affordable Prices
www.website12.com

Cuba Travel

Plan your trip with Twindex® and
Reduce the Risk of Hepatitis A & B
www.website13.com

Cuba Vacations 30% Off

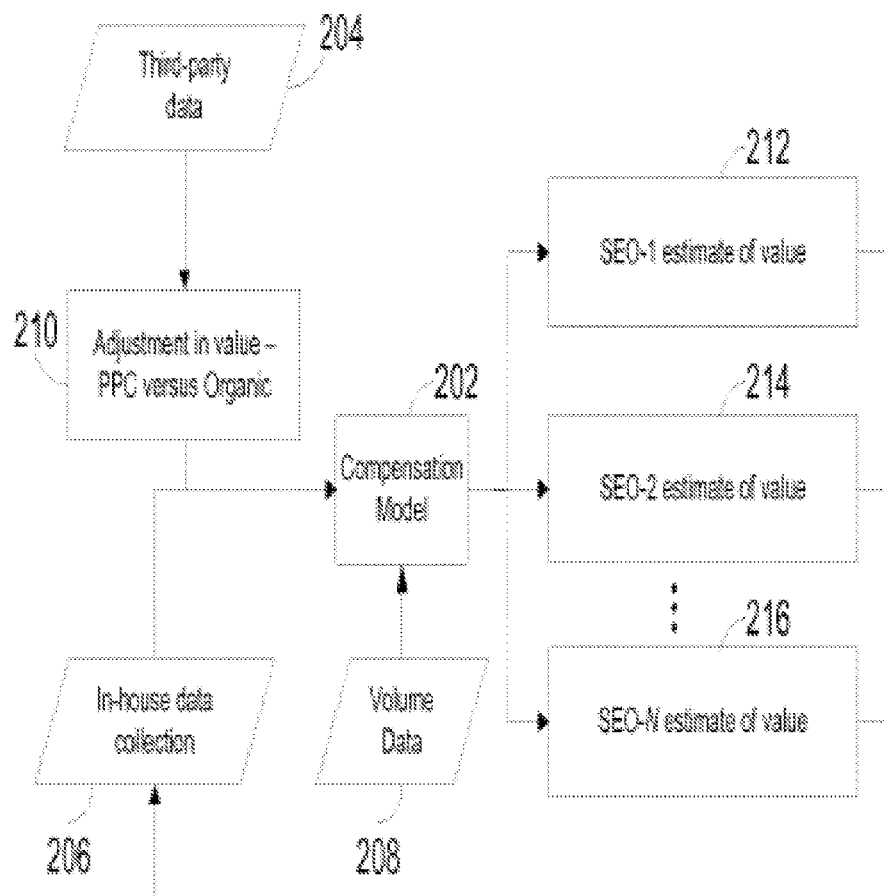
All Inclusive Sale to Cuba
Book Online in Canadian Dollars
www.website14.com

Travel

Sponsored
Search
Results

Organic
Search
Results

FIG. 1
PRIOR ART



Estimating Value of Placement
in an Organic SERP

FIG. 2

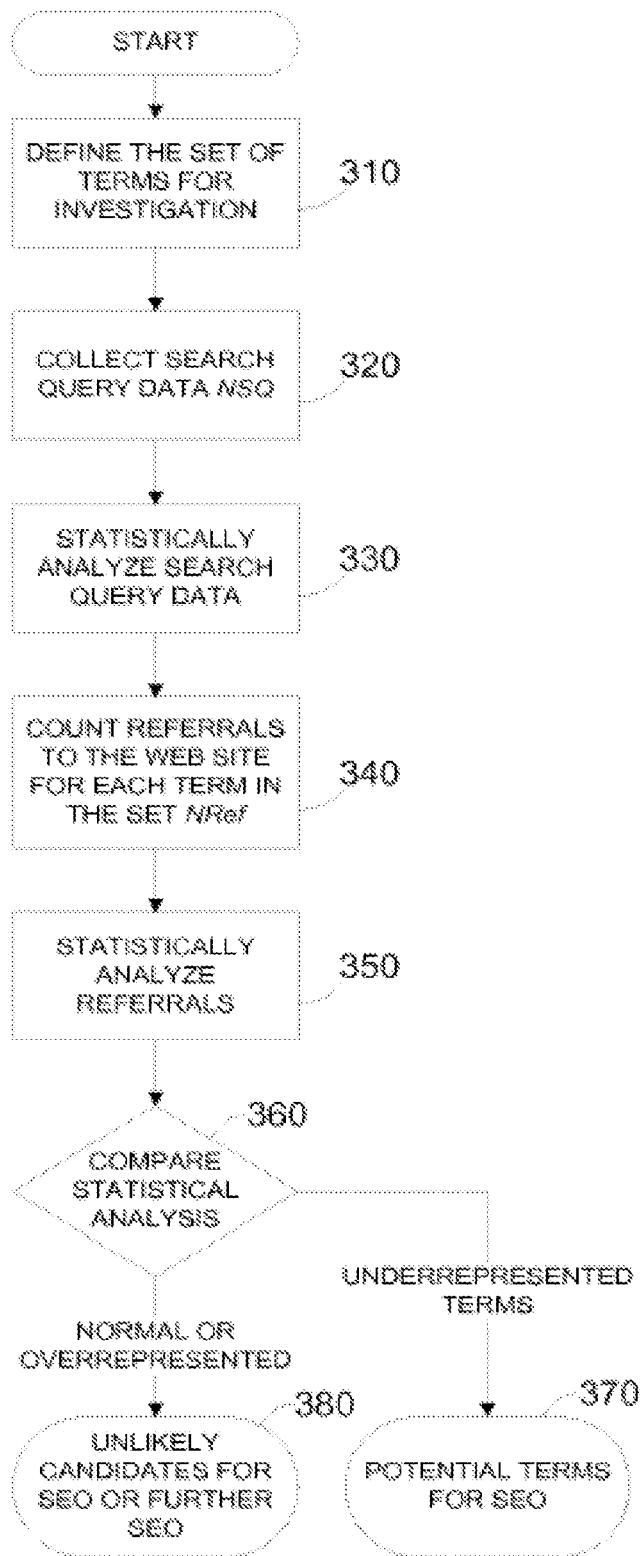


FIG. 3

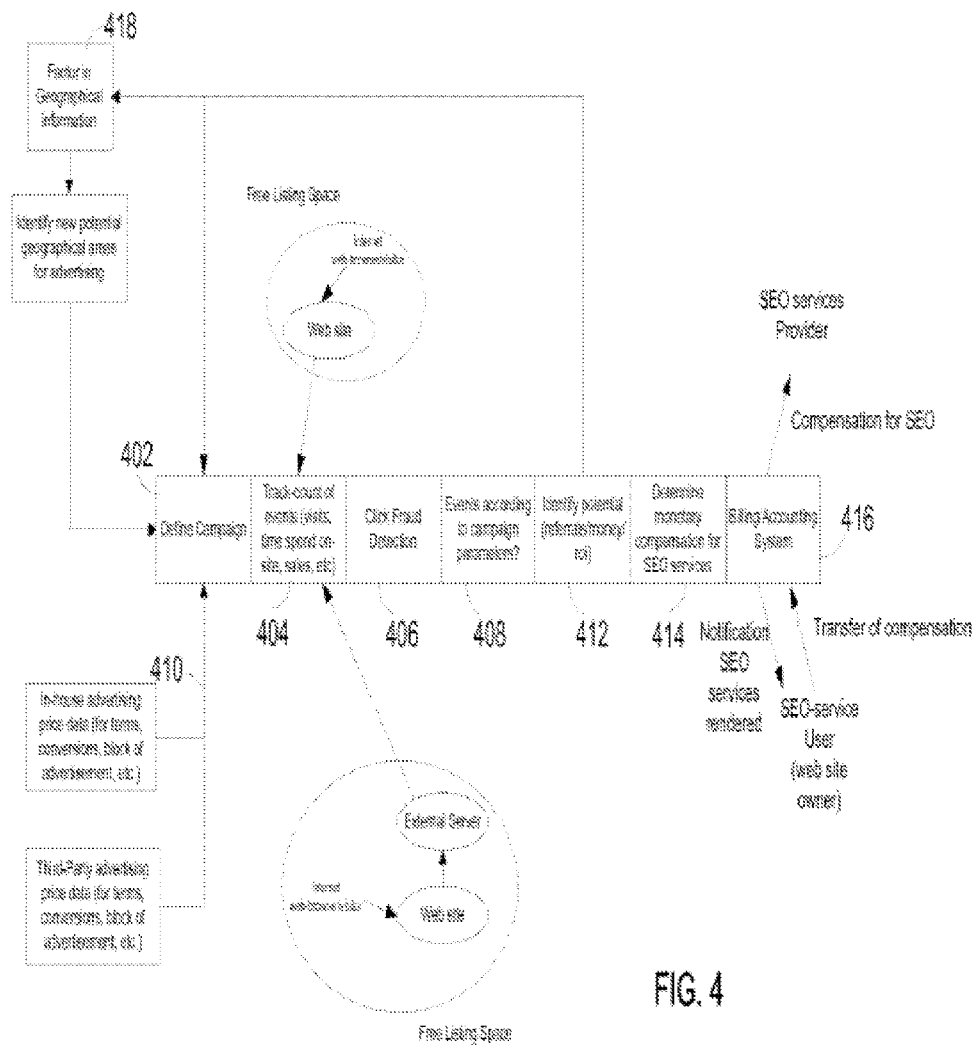


FIG. 4

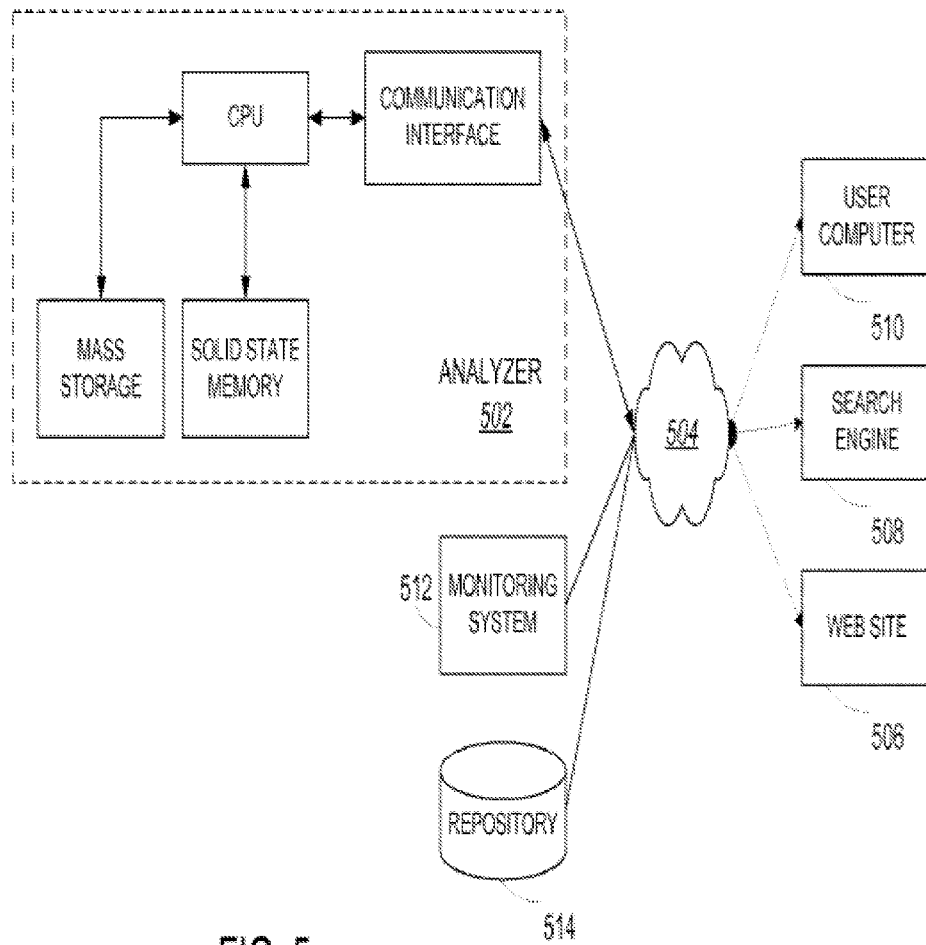


FIG. 5

SEARCH ENGINE OPTIMIZATION PERFORMANCE VALUATION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a divisional of U.S. patent application Ser. No. 12/555,224, filed Sep. 8, 2009, now U.S. Pat. No. 8,364,529, issued Jan. 29, 2013, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. 61/094,821, filed Sep. 5, 2008, each of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

The invention relates to online marketing, and in particular, to automation for monitoring and accounting in connection with search engine optimization (SEO) marketing activities.

BACKGROUND

Description of the Related Art

A search engine is a program that assists users to search for information on the Internet. When an Internet user enters a search query into an interface for a search engine, the search engine examines its index and provides a listing of matching documents. The usefulness of a search engine depends on the relevance of the result set it generates and shows to the web user. While there may be many documents that include a particular word or phrase, some documents may be more relevant, popular, or authoritative than others.

Some search engines provide search-related advertisements with regular search engine results in a search engine results page (SERP). A natural result set that should not be influenced by payment for listing is known as an organic search result and occupies a space within the total result set. A result set influenced by payment for listing is known as a sponsored search result. FIG. 1 illustrates a conventional search engine results page (SERP) which shows both organic and sponsored results.

One way to promote a web site is by purchasing advertising from advertising networks or publishers, such as on a pay per click (PPC) basis. Typically, advertising is selected based on an auction for the advertising space. While an advertising network may offer a defined price for advertising, such as an amount per click, it is more common for prices to be set through a bidding or auction arrangement.

Another way to direct traffic to a web site is via search engine optimization (SEO). A goal of SEO is to improve the visibility of a web site by making listings for the web site appear more frequently and more prominently in the free (organic) portion of a search engine results listing. Many strategies exist for performing SEO.

While SEO can be performed by operators of a web site itself, SEO services are typically performed by consultants or companies having specific SEO expertise. These SEO practitioners contract with a web site's operators to improve the ranking of links to the web site in a search engine's organic listing. For example, SEO practitioners can be compensated on a fixed fee basis, which can be supplemented with bonuses tied to specific results, such as (1) how high links to the web site are placed on the organic listing, (2) increases in Internet traffic (for example, page hits), or (3) increases in sales as a result of better (higher) Internet traffic, or to combinations of the same.

SUMMARY

A way to promote a web site is via search engine optimization (SEO). Traditionally, SEO practitioners have charged a fixed fee for performing SEO services. An analyzer can determine a first probability that a search query for a campaign term will result in a referral from a search engine, determine second probabilities that are associated with a plurality of particular positions in a search engine results page, combine the search query volume information, the first probability, and the second probability for the position with a monetary value to generate an organic price, combine the organic prices for the plurality of positions to value the SEO services for the campaign term. The analyzer evaluates the effectiveness of SEO activities and can generate a volume-based value for SEO services, which can be used as a reference for providing SEO services.

In addition, disclosed techniques can also be used to identify candidate terms for SEO for future services. An analyzer computes rates of referrals for each term in a set of terms. The set of terms comprises, for example, terms that are used in search engine queries to access the site. The analyzer identifies terms in the set that are underrepresented in terms of rates of referral relative to one or more other terms in the set, and these underrepresented terms are identified as being suitable for SEO.

BRIEF DESCRIPTION OF THE DRAWINGS

These drawings and the associated description herein are provided to illustrate specific embodiments of the invention and are not intended to be limiting.

FIG. 1 illustrates an example of a search engine result page (SERP).

FIG. 2 illustrates inputs and outputs for an estimation model for valuing organic referrals.

FIG. 3 is a flowchart generally illustrating a process for identifying new campaign terms (keywords).

FIG. 4 is a data flow diagram generally illustrating interaction with web sites, with SEO service provides, and the like.

FIG. 5 illustrates an analyzer 502 can be used to implement an embodiment of the invention.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Although particular embodiments are described herein, other embodiments of the invention, including embodiments that do not provide all of the benefits and features set forth therein, will be apparent to those of ordinary skill in the art. In addition, while illustrated in the context of web sites, the principles and advantages described herein are applicable to other types of sites and interactive systems that are accessible on a network.

Traditionally, SEO practitioners have charged a fixed fee for performing SEO services. While positions of listings (rankings) within an organic search result are not sold individually as PPC advertisements can be, these positions nonetheless have value. Techniques quantitatively monitor and value these rankings, and can generate a volume-based value for SEO services, which can be used as a reference for providing SEO services. These techniques can be implemented by software executed by computer hardware. All of the methods and processes described may be embodied in, and fully automated via, software code modules executed by computer hardware, such as one or more general purpose computers. The code modules may be stored on any type of physical

computer-readable medium or computer storage device. Some or all of the methods may alternatively be embodied in specialized computer hardware. The results of the disclosed methods (e.g., generated data) may be stored persistently on any type of physical computer data storage media or system. Instructions for the software can be stored in a tangible, computer-readable medium. Applicable computer systems include file servers, web servers, database software, network monitoring equipment, and the like.

One embodiment is a technique that automatically determines valuation for search engine optimization (SEO) services. Typically, particular phrases or keywords are selected as campaign terms for SEO by the web site operator. A SEO practitioner is then hired to perform SEO for the campaign terms. These campaign terms relate to query phrases used by Internet users when searching for information via a search engine tool. When, due to SEO activities, the campaign terms are identified as having been used by Internet users as search queries in associated search engine referrals, SEO activities can be valued.

To determine a valuation for SEO services, the illustrated embodiment of the invention uses "in-house," which can be proprietary data related to SEO, and third-party data related to pay per click (PPC) advertising. In one embodiment, the in-house or proprietary data comprises at least previously collected search engine referral data and pricing data for the same or related terms, and can be collected algorithmically or manually. One embodiment of the invention permits SEO practitioners to establish an intrinsic value for campaign terms, which can be used as a reference for SEO optimization services. For example, averages of these values for SEO services can be provided as values for similar terms; these values can vary according to performance metrics including: particular positions of links in listing, volumes of searches for campaign terms, and long tail phrases for campaign terms. To determine value for campaign terms, one embodiment of the invention uses third-party auction networks prices, for example, PPC pricing; however, these third-party prices are related to sponsored listings.

Valuation of Campaign Terms in Organic Searches

FIG. 2 illustrates inputs and outputs for a compensation model 202 for valuing organic referrals. The compensation model 202 can use third-party data 204, such as price information for PPC advertisements, in-house data 206, and volume data 208.

The third-party data 204 can include, for example, PPC price data and PPC volume data. The third-party data 204 can further be adjusted 210 to account for the difference in usage of selection of sponsored links versus organic links within a SERF. For example, research indicates that informational searches are more common for organic listings than in sponsored listings. Many web site operators are interested only in specific transactional operations, such as sales, and thus, these operators will typically not value queries from organic search results equally as sponsored listings. Thus, sponsored listing prices should be adjusted to reflect these differences in value, such as: (1) differences in the fractions of transactional operations; (2) differences in the value of transactional operations; (3) the position of the listing in the organic space; and (4) the human perception that organic referrals are not as valuable as purchased (for example, PPC) referrals.

The in-house data 206 can include, for example, information regarding historical SEO service pricing for campaign terms, ranking performance, historical search activity volume, demographic data to identify potentially similar or like user profiles, global search activity volumes and behaviors, and referral volume behavior and patterns from non-search

activity, such as hyperlink referrals wherein the hyperlink has anchor text (for example, anchor text) that is relevant to a particular term, and thus the activity from that link can be valued against outcomes from referrals. An outcome can be any monetary or monetary equivalent value which may be assigned to an action (such as the value of a conversion) which occurred as a result of a visitor who arrived via that referral. The volume data 208 can include, for example, information related to the volume of searches for particular terms, the rates of selection of links in particular positions, and the like.

One embodiment of the compensation model 202 will be described in greater detail later in connection with Equation 1. The compensation model 202 can generate a reference for a price for SEO services for a particular campaign term. For example, the value for SEO that results in the topmost ranking organic search result (SEO-1) 212 can be estimated by the compensation model 202. Similarly, other rankings such as the second highest ranking organic search result (SEO-2) 214 and an arbitrary N-th ranking SEO-N 216 can be computed. In an alternative embodiment, rather than represent organic search result ranking, blocks SEO-x in FIG. 2 can represent different SEO practitioners, and the compensation model 202 can be used to track the SEO performance of the various SEO practitioners. In an alternative embodiment, the compensation model 202 ranks these different SEO practitioners based on the relative performance of their SEO efforts, and generates a weighting for the different SEO practitioners.

Thus, SEO-x can represent at least one of: (1) an organic search result rank; (2) a particular SEO practitioner; or (3) a weight an SEO practitioner has among all SEO practitioners (rank). Any or a combination of these three can be used as inputs for the generation of an intrinsic value for campaign terms.

An organic price for placing a campaign term in a particular position (ranking) can depend on multiple factors, such as, but not limited to: (1) the number of search queries for the campaign term performed on a given network; (2) the probability that the volume (number of search queries) will result into actual referrals; and (3) the probability that such referrals take place from the desired position (ranking). When data is available, such as after SEO has been performed and data has been collected, the foregoing probabilities can be replaced with fractions based on actual data. These factors can be used to calculate the number of potential events. These events can include, but are not limited to referrals, clicks, visits, conversions, in which an SEO campaign term is used by the user in the search query, and wherein the campaign term present within the specified position or ranges of positions on the search engine results page (SERP).

The calculated number of events can be multiplied by a price per event. A price for similar events can be obtained from in house price data, which typically includes previously-collected search engine referral data and prices, whether collected algorithmically or via manual input for organic placement and/or from third-party data for placement on the paid-space.

More than one in-house source and third-party source of price data can be used as inputs. The in-house or proprietary data can change over time and can also vary seasonally, and thus, data covering vary sets of time can be selected. In one embodiment, the data is seasonally adjusted. For example, multiple sources of PPC data can be compiled from various search engines. In one example, price data from each source can be weighted according to pre-defined rules. For example, a volume-weighted price-average can be computed from multiple sources of PPC data, and the weighted average can be

used to generate values for campaign terms. Pre-defined rules can also include weights that represent a perceived performance/usefulness of the various search engines from which PPC data is obtained. As is well known, some search engines produce more ‘meaningful’ or useful results than others. While the perception of meaningfulness is qualitative, a weight can be assigned to each search engine to adjust the PPC data of the search engines to account for the differences in usefulness. Pre-defined rules can also include weights representing the fact that organic referrals tend to be more information seeking than referrals from paid listing. SEO values can be updated as data is collected based on actual usage, and the collected data can be kept in the in-house data collection 206.

Equation 1 expresses an example of a mathematical model for estimating the value of placement of a link in a particular position Y in the organic space of a search engine results page. In the illustrated embodiment, the mathematical model uses both in-house and third-party price data as inputs. In an alternative embodiment, either in-house or third-party cost data suffices to estimate organic prices.

$$OP(x,y)=(S_x)(P_{R_y})(\Sigma(W)(IP)+\Sigma(M)(TP)) \quad (\text{Eq. 1})$$

In Equation 1, $OP(x,y)$ represents the organic price for placement of campaign term x in position y (position y is a particular position within a search engine results page; S_x represents the search query volume for campaign term x; P_{R_y} represents the fraction (when known, probability otherwise) of the search query phrase volume that converts into referrals; P_{R_y} represents the fraction (when known, probability otherwise) of referrals listed in position y; IP represents one or more in-house sources of price data, W represents one or more weights denoting the importance (for example, rank) of each one of the multitude of in-house sources of price data; $\Sigma(W)$ (IP) represents the sum of the one or more in-house sources of price data each source properly weighted; TP represents one or more third-party sources of price data, M represents one or more weights denoting the importance (for example, rank) of each of the third party sources of price data; $\Sigma(M)(TP)$ represents the sum of the one or more third-party sources of price data with each source properly weighted. In the illustrated embodiment, the formula expressed in Equation 1 is applied to a particular search engine to determine an organic price for placement on that search engine. In an alternative embodiment, such as, for example, for use with a relatively new search engine with little to no prior history, a combined, such as an average, organic price across multiple search engines can be generated. The data considered in Equation 1 can also be restricted to geographical regions, to specified time periods, and the like. In one embodiment, the weighted values $\Sigma(W)(IP)$ and $\Sigma(M)(TP)$ also vary by position y. The results over various campaign terms x and positions y can be summed.

Compensation models 202 can be combined with thresholds and/or limits to activate/de-activate the compensation model 202. For example, the compensation model 202 can be supplemented with a condition of not charging for events until the total number of events has reached a minimum threshold. In another example, the compensation calculated by the compensation model 202 can reach a cap when the compensation has reached an upper limit.

In an alternative embodiment, pricing for events other than campaign terms are modeled, such as pricing for external hyperlinks (for pages other than SERP) and conversions that are the result of visit due to SEO efforts. When these additional events are the result of SEO efforts, the events should be assigned values so that SEO efforts are properly compen-

sated. In one example, a SEO practitioner and a site can specify that particular pages are included in a SEO campaign, or that all pages within the site may be included. In addition, offsite pages (external hyperlinks) can also be similarly identified. Typically, the arrangement between the SEO and the site is handled manually. Referring hyperlinks to the web site for which SEO is being performed are commonly recruited for at least two purposes: to expose the web site to additional visitors via the hyperlinks; and to increase a web site’s importance (search engine algorithms can use the prevalence of links to documents/web sites to value the importance of a document or a web site). By procuring good quality links, a SEO practitioner attracts both new customers and also increases the exposure of the web site within the search engine result pages. Some link referrals and some search engine referrals will result in conversions; and if these referrals are the result of SEO efforts their corresponding conversions should be assigned values so SEO activities are compensated.

A variety of techniques can be used for assessing organic internet advertising performance. For example, the processes described in commonly-owned U.S. patent application Ser. No. 12/143,387, filed Jun. 20, 2008, titled “INTERNET MARKETING VERIFICATION TOOL,” the disclosure of which is hereby incorporated by reference herein can be used to collect http referrers (referring hyperlinks or refer data) used to access a SEO optimized web site and to evaluate the collected refer data against the campaign parameters for assessment of the Internet advertising. See, for example, FIG. 4 and the accompanying description of U.S. patent application Ser. No. 12/143,387. Campaign parameters include geographical origins of refer data, language preference of web user/visitor, and landing pages for refer data.

Identification of New Campaign Terms (Keywords)

To drive even more volume to the pages of a web site, SEO can be performed on additional terms. For example, there can be more than one way to describe a topic, there can be more than one way to spell a word, and the like. Preferably, all relevant terms are optimized by SEO. However, oftentimes, there are terms that are unknown or otherwise not recognized when SEO is being performed. This can happen frequently with, for example, new lingo, alternative spellings, or current events. These new terms can be used as new campaign terms or keywords for SEO, further enhancing SEO efforts for the web site for which SEO is being performed.

In the context of online marketing, a web site operator would prefer to have its search result listings in a relatively high position of a SERP, such as, for example, the first position of the first page of a SERP. However, some search terms are more competitive than others and achieving a desired position for a competitive search term may not be practical for a particular web site. However, there may be other search terms from which a web site receives hits (or can potentially receive hits) that are not as competitive as the most competitive search terms, and SEO can be performed for these other search terms to achieve higher traffic for the web site.

FIG. 3 illustrates a process that can be used to test and/or identify new campaign terms (keywords). In this context, “new” relates to the identification of terms for SEO efforts other than existing campaign terms, which, after SEO can generate statistically meaningful additional traffic to the web site. The process can be used to identify search terms to be used for campaign terms, and can alternatively be used to determine whether it is worth the expense to perform SEO for proposed campaign terms.

The illustrated process can be modified in a variety of ways. For example, in another embodiment, various portions

of the illustrated process can be combined, can be rearranged in an alternate sequence, can be removed, or the like. At the start of the process, it is assumed that data for analysis has already been collected.

The process begins at a state **310** by defining the set of terms for investigation by the process. The set of terms are those terms that are at least worthy of investigation for possible SEO for the particular web site, and thus, should be over inclusive. These terms can be words or phrases. These terms can be, for example, terms that were found to have been used by users as search queries to access web pages of the web site, terms derived from analysis of web pages of the web site for which SEO services are being considered, associated PPC terms, and any other term deemed worthy of investigation. The process then analyzes these terms to determine whether SEO of these terms would be beneficial as described below.

The process advances from the state **310** to a state **320** to collect search query data. For example, for each term in the set of terms, the process can determine the corresponding search query volume, that is, how many times users used the particular term as a search query term. The count can be limited to one or more selected search engines, to particular countries or other geographic limitations, to particular time or date ranges, and the like. In one embodiment, the count is determined by retrieving the query volume information from a variety of search engines and combining the retrieved information. In the tables that follow, this count corresponds to the variable NSQ.

An appropriate time or date range will vary depending upon the popularity of a web site. For example, a high-volume web site may generate enough traffic for analysis in a relatively short period of time. However, a relatively low-volume web site may need to have data collected over a relatively lengthy time period, such as over 10 days, to collect enough data to be analyzed meaningfully. The process advances from the state **320** to a state **330**.

In the state **330**, the process statistically analyzes the search query data for each term of the set. In a simple example, the process can compute the fraction of the total search query volume for each term in the set, that is, the volume of search queries for each particular term of the set divided by the total volume of search queries in the set. This statistical analysis establishes a point of reference for referrals. The process advances from the state **330** to a state **340**.

In the state **340**, the process counts the actual referrals to the web site for each term in the set. In the tables that follow, this count corresponds to the variable NRef. The prior constraints used in the state **320**, such as geographical constraints, time or date range constraints, and the like, should again be applicable in the state **340**. The process advances from the state **340** to a state **350**.

In the state **350**, the process analyzes the actual referrals. In one example, the process can compute the fraction of the total actual referral volume for each term in the set, that is, the volume of referrals for each particular term of the set divided by the total volume of referrals in the set. This statistical analysis establishes a point of reference for referrals. The process advances from the state **350** to a decision block **360**.

In the decision block **360**, the process compares the statistical analysis of each of the terms in the set. In one example, the process compares the fraction of the total actual referral volume for a particular keyword versus the fraction of the total search query volume (baseline reference) for said keyword. The comparison is used to determine whether the term is a good candidate for a campaign term SEO. For example, if the fraction of total actual referral volume is lower than a pre-defined amount of the fraction of the total search query

volume (baseline) then said term is identified as underrepresented (for example, 30% versus 50%).

In another example, the process compares the ranking of the fraction of the total actual referral volume for a particular keyword (term) versus the ranking of the fraction of the total search query volume (baseline reference) for the particular keyword (term). For example, if the ranking of the fraction of the total actual referral volume is lower than a predefined amount of the ranking of the fraction of the total search query volume (baseline) then the keyword (term) is identified as underrepresented (for example, rank **2** versus rank **5**).

In another example, the process compares the ranking of the fraction of the total actual referral volume for all keywords versus the ranking of the fraction of the total search query volume (baseline reference) for all keywords. For example if the set of fractions of the total actual referral volume for all keywords is statistically different from the set of fractions of the total search query volume for all keywords then a delta in fractions to the set of fractions of the total actual referral volume is estimated so that the new fractions are similar (or identical) to the fractions of the total search query volume. For example, for each term in the set, the process computes a value for term volume TV as the maximum of the count of referrals NRef or the count of the search query volume NSQ. The maximum of the two is used to cap a later computed value for term ratio TR, as there can be discrepancies when the number of referrals from sources outside of search engines is relatively large such that there will be more referrals than search queries (sometimes search query volumes are estimated from samples of the universe of search queries; if samples are not representative then these may have no correlation with a TRue count (no sampling) of the universe of search queries) or when there are problem with data collection, or the like. The process then computes a value for the term ratio TR as the count of referrals NRef divided by the term volume TV. The process compares the term ratio TR of each of the terms to a baseline reference to determine whether the term is a good candidate for a campaign term SEO. In one embodiment, the baseline reference is the best performing term in the set of terms, and terms below a predetermined percentage of the baseline reference value are identified for SEO.

The terms having a referral rate below a predefined amount of the reference baseline value are identified as underrepresented and potential terms **370** for SEO. The predefined amount can vary widely based on a variety of factors, such as on the overall value of SEO for the web site. For example, underrepresented can be determined to be about 10% or more below the reference baseline value, but other applicable values will be readily determined by one of ordinary skill in the art. The terms that already have a similar referral rate to the baseline or are already overrepresented are unlikely **380** to make good candidates for SEO.

In one embodiment, the process calculates a hypothetical amount of volume that could have been driven to the web site had selected terms in the set been optimized as campaign terms for SEO. For example, the process can calculate the difference in rate between an expected rate and the actual rate, and multiply that difference in rate with the number of search queries NSQ for those terms. In addition, conversion rates can be used to determine an amount of revenue that would have been generated. These tools permit a web site operator and/or SEO practitioner to determine whether it is worth the effort to perform SEO on selected terms.

Table 1 illustrates an example of computations using the process of FIG. 3. Sample keywords labeled "A" to "D" are used. In this example, keyword "D" has the largest term ratio

TR of 0.75 and is used as a baseline reference. The “Expected Traffic” computes the expected number of referrals on the site of interest if these terms were to be included in SEO. The “Potential Benefit” column is the difference between the expected traffic and the actual number of referrals NRef, and the column indicates the additional volume that the web site can potentially receive per term if the terms are included in SEO efforts.

TABLE I

Term	Search Query (NSQ)	Number Referrals (NRef)	Term Volume (TV)	Term Ratio (TR)	Expectation Volume	Potential Benefit
A	100	40	$\max(40, 100) = 100$	$40/100 = 0.4$	$100 \times 0.75 = 75$	$\max(0, 75 - 40) = 35$
B	15	10	$\max(10, 15) = 15$	$10/15 = 0.66$	$15 \times 0.75 = 11.25$	$\max(0, 11.25 - 20) = 2.5$
C	30	20	$\max(20, 30) = 30$	$20/30 = 0.66$	$30 \times 0.75 = 22.5$	$\max(0, 22.5 - 20) = 2.5$
D	40	30	$\max(30, 40) = 40$	$30/40 = 0.75$	$40 \times 0.75 = 30$	$\max(0, 30 - 30) = 0$

Table II illustrates that the percentage distribution of the “Expected Volume” should be about the same (identical in the example) to that of the “Search Query” volume.

TABLE II

Term	Search Query (NSQ)	Distribution Search Query	Expectation Volume (referrals)	Distribution ideal traffic
A	100	$54\% = 100/185$	75	$54\% = 75/138.75$
B	15	$8.1\% = 15/185$	11.25	$8.1\% = 11.25/138.75$
C	30	$16.2\% = 30/185$	22.5	$16.2\% = 22.5/138.75$
D	40	$21.6\% = 40/185$	30	$21.6\% = 30/138.75$
Total	185	N/A	138.75	N/A

FIG. 4 is a data flow diagram generally illustrating interaction with web sites, with SEO service provides, and the like. Campaign events **402** are defined (referrals; hyperlinks; conversions; etc) to track. SEO results are determined **404** by, for example, counting/tracking pre-defined events (for example, visits, actions, sales). Fraudulent events are identified and filtered **406** (for example, high traffic but no sales). Events are verified **408** that they match pre-defined parameters (for example, sales from the United States). Values are assigned **410** to events. Sources of revenue are identified **412** (for example, potential referrals/money/return-on-investment). Monetary compensation is determined **414** for SEO services provided or to be provided. A billing I accounting mechanism **416** provides billing to web site operator based on the value of work delivered, and to the search marketer based on performance of work delivered. A feedback loop mechanism **418** modifies a campaign based on sources of revenue (potential). After terms with relatively good potential have been identified, these identified terms can be forwarded to SEO practitioners for optimization.

FIG. 5 illustrates an analyzer **502** can be used to implement an embodiment of the invention. For example, the analyzer **502** can be configured to implement the process described earlier in connection with FIG. 2 or the process described earlier in connection with FIG. 3. The various systems illustrated in FIG. 5 can be implemented by computer hardware that executes software instructions.

FIG. 5 illustrates an the analyzer **502**, a network **504**, a web site **506** for SEO, a search engine **508**, a user computer **510**, a monitoring system **512**, and a data repository **514**. The net-

work **504** can correspond to local area networks and to wide area networks, such as the Internet. The web site **506** represents a web site that is being optimized by SEO practitioners or is considering using SEQ. The search engine **508** provides search engine results pages (SERP) to the user computer **510**. The monitoring system **512** and data repository **514** or data store can be used to collect http referrers (referring hyperlinks or refer data) used to access a SEO-optimized web site and to

evaluate the collected refer data against the campaign parameters for assessment of the Internet advertising.

Various embodiments have been described above. Although described with reference to these specific embodiments, the descriptions are intended to be illustrative and are not intended to be limiting. Various modifications and applications may occur to those skilled in the art.

What is claimed:

1. A method of identifying terms suitable for search engine optimization (SEO) for a site, the method comprising:

determining a volume of search queries for each term in a set of terms, wherein the set of terms comprises at least a plurality of terms used in search engine queries to access the site;

counting referrals to the site for each term in the set of terms;

computing a rate of referral for each term in the set of terms as a ratio of the counted referrals to the volume of search queries;

computing an expectation volume for each term in the set of terms based on the highest rate of referral of the terms in the set of terms and the volume of search queries for each term in the set of terms; and

identifying one or more terms in the set of terms that are underrepresented relative to one or more other terms in the set of terms based on the expectation volume for each term in the set of terms, wherein the one or more underrepresented terms are identified as being suitable for SEO;

wherein the determining, the counting, and the computing are performed, at least in part, by one or more computing devices.

2. The method of claim 1, wherein identifying the one or more terms in the set of terms that are underrepresented comprises:

identifying that the one or more underrepresented terms have rates of referral outside of a selected statistical measure of a baseline reference, wherein the baseline reference is based on a highest rate of referral for a term in the set of terms.

3. The method of claim 2, wherein the selected statistical measure is one standard deviation.

4. The method of claim 1, further comprising: selecting terms for inclusion in the set of terms based on a specified time period or geographic region.

11

5. The method of claim 1, further comprising:
generating an estimate of additional referrals to the site due
to additional referrals that would be generated if the SEO
activities were performed for the one or more underrep-
resented terms; and

generating at least one of a revenue estimate or a profit
estimate based on the additional referrals.

6. An apparatus for identifying terms suitable for search
engine optimization (SEO) for a site, the apparatus compris-
ing:

a data store configured to store information related to
search terms and referrals to the site;

an analyzer comprising one or more computing devices,
said analyzer configured to:

compute a rate of referral for each term in a set of terms
as a ratio of counted referrals to a volume of search
queries, wherein the set of terms comprises at least a
plurality of terms used in search engine queries to
access the site, wherein the counted referrals comprise
referrals to the site for each term in the set of
terms, and wherein the volume of search queries com-
prises a volume of search queries for each term in the
set of terms;

compute an expectation volume for each term in the set
of terms based on the highest rate of referral of the
terms in the set of terms and the volume of search
queries for each term in the set of terms; and

identify one or more terms in the set of terms that are
underrepresented relative to one or more other terms
in the set of terms based on the expectation volume for
each term in the set of terms, wherein the one or more
underrepresented terms are identified as being suit-
able for SEO via statistical analysis.

7. The apparatus of claim 6, wherein the analyzer is con-
figured to identify the one or more terms in the set of terms
that are underrepresented by identifying that the one or more
underrepresented terms have rates of referral outside of a
selected statistical measure of a baseline reference, wherein
the baseline reference is based on the highest rate of referral
for a term in the set of terms.

8. The apparatus of claim 7, wherein the selected statistical
measure is one standard deviation.

9. The apparatus of claim 6, wherein the terms in the set of
terms are selected for inclusion in the set of terms based on a
specified time period or geographic region.

10. The apparatus of claim 6, wherein the analyzer is fur-
ther configured to:

generate an estimate of additional referrals to the site due to
additional referrals that would be generated if the SEO
activities were performed for the one or more underrep-
resented terms; and

generate at least one of a revenue estimate or a profit
estimate based on the additional referrals.

11. A non-transitory computer readable medium having
instructions embodied thereon for identifying terms suitable
for search engine optimization (SEO) for a site, wherein the
instructions, in response to being executed by a computing
device, cause the computing device to:

determine a volume of search queries for each term in a set
of terms, wherein the set of terms comprises at least a
plurality of terms used in search engine queries to access
the site;

count referrals to the site for each term in the set of terms;
compute a rate of referral for each term in the set of terms
as a ratio of the counted referrals to the volume of search
queries;

12

compute an expectation volume for each term in the set of
terms based on the highest rate of referral of the terms in
the set of terms and the volume of search queries for each
term in the set of terms; and

identify one or more terms in the set of terms that are
underrepresented relative to one or more other terms in
the set of terms based on the expectation volume for each
term in the set of terms, wherein the one or more under-
represented terms are identified as being suitable for
SEO.

12. The non-transitory computer readable medium of claim
11, wherein the instructions that cause the computing device
to identify the one or more terms in the set of terms that are
underrepresented comprise instructions that, in response to
being executed by the computing device, cause the computing
device to identify that the one or more underrepresented terms
have rates of referral outside of a selected statistical measure
of a baseline reference, wherein the baseline reference is
based on a highest rate of referral for a term in the set of terms.

13. The non-transitory computer readable medium of claim
12, wherein the selected statistical measure is one standard
deviation.

14. The non-transitory computer readable medium of claim
11, wherein the instructions, in response to being executed by
the computing device, further cause the computing device to:
select terms for inclusion in the set of terms based on a
specified time period or geographic region.

15. The non-transitory computer readable medium of claim
11, wherein the instructions, in response to being executed by
the computing device, further cause the computing device to:
generate an estimate of additional referrals to the site due to
additional referrals that would be generated if the SEO
activities were performed for the one or more underrep-
resented terms; and

generate at least one of a revenue estimate or a profit
estimate based on the additional referrals.

16. An apparatus for identifying terms suitable for search
engine optimization (SEO) for a site, the apparatus compris-
ing:

means for determining a volume of search queries for each
term in a set of terms, wherein the set of terms comprises
at least a plurality of terms used in search engine queries
to access the site;

means for counting referrals to the site for each term in the
set of terms;

means for computing a rate of referral for each term in the
set of terms as a ratio of the counted referrals to the
volume of search queries;

means for computing an expectation volume for each term
in the set of terms based on the highest rate of referral of
the terms in the set of terms and the volume of search
queries for each term in the set of terms; and

means for identifying one or more terms in the set of terms
that are underrepresented relative to one or more other
terms in the set of terms based on the expectation volume
for each term in the set of terms, wherein the one or more
underrepresented terms are identified as being suitable
for SEO;

wherein the determining, the counting, and the computing
are performed, at least in part, by one or more computing
devices.

17. The apparatus of claim 16, wherein the means for
identifying the one or more terms in the set of terms is con-
figured to identify that the one or more underrepresented
terms have rates of referral outside of a selected statistical

measure of a baseline reference, wherein the baseline reference is based on a highest rate of referral for a term in the set of terms.

18. The apparatus of claim **17**, wherein the selected statistical measure is one standard deviation. 5

19. The apparatus of claim **16**, further comprising:
means for selecting terms for inclusion in the set of terms
based on a specified time period or geographic region.

20. The apparatus of claim **16**, further comprising:
means for generating an estimate of additional referrals to 10
the site due to additional referrals that would be generated if the SEO activities were performed for the one or more underrepresented terms; and
means for generating at least one of a revenue estimate or
a profit estimate based on the additional referrals. 15

* * * * *